UNIVERSITETET I OSLO

Det matematisk-naturvitenskapelige fakultet

Eksamen i:	ST102 — Videregående kurs i statistikk.
Eksamensdag:	Mandag 29. mai 2000.
Tid for eksamen:	09.00 – 15.00.
Oppgavesettet er på 4 sider.	
Vedlegg:	1) 53 observasjoner om effekt av blodtrykksregulerende middel. 2) Spredningsplott 3) Tabell over normal- og t-fordelingen.
Tillatte hjelpemidler:	Formelsamlinger i ST101 og ST102, Rottmanns formelsamling, lommekalkulator.

Kontroller at oppgavesettet er komplett
før du begynner å besvare spørsmålene.

Oppgave 1.

La $X_1, \ldots, X_n$ være uavhengige eksponensielt fordelte variable med tetthet

$$f(x) = \frac{1}{\tau}\exp(-x/\tau) \qquad x > 0\ , \quad \tau > 0\ .$$

a) Bestem momentestimatoren til τ.

b) Sett opp likelihoodfunksjonen og finn sannsynlighetsestimatoren $\widehat{\tau}$ til τ.

c) Vis at $2\sum_{i=1}^{n} X_i/\tau$ er χ^2 fordelt og finn fordelingen til $\widehat{\tau}$. Angi forventning og varians.

d) Bruk sentralgrenseteoremet til å finne den tilnærmede fordelingen til $\widehat{\tau}$ når n er stor.

(Fortsettes side 2.)

e) Forklar sammenhengen mellom resultatet du fant i punkt d) og det generelle resultatet for den tilnærmede fordelingen til sannsynlighetsestimatoren når n er stor.

Oppgave 2.

Vi skal i denne oppgaven se på $Q-Q$ plot for den dobbelteksponensielle fordelingen. Den har tetthet gitt ved

$$f(x) = \tfrac{1}{2}e^{-|x|} \qquad -\infty < x < \infty\ .$$

Som tettheten til en standard normalfordelt tilfeldig variabel er tettheten symmetrisk rundt null. Tettheten avtar imidlertid saktere mot null for $x \to \pm\infty$.

a) Finn den kumulative fordelingsfunksjonen F, og vis at den inverse funksjonen F^{-1} er gitt ved

$$F^{-1}(z) = \begin{cases} \log(2z) & 0 < z < \frac{1}{2} \\ -\log(2(1-z)) & \frac{1}{2} \le z < 1 \end{cases}$$

b) La U være en uniformt fordelt variabel på intervallet $[0,1]$, d.v.s. fordelingen til U har tetthet

$$g(u) = \begin{cases} 1 & 0 \le u \le 1 \\ 0 & \text{ellers}\ . \end{cases}$$

Vis at den tilfeldige variabelen $F^{-1}(U)$ er dobbelteksponensielt fordelt.

c) Anta at $U_1, \dots, U_n$ er uavhengige uniformt fordelt på $[0,1]$ og la $U_{(1)} \le \dots \le U_{(n)}$ være ordningsobservatoren. La $V_k = U_{(k)}$ være k'te element i ordningsobservatoren. Da har V_k tetthet, $k = 1, \dots, n$

$$h(v) = \begin{cases} \frac{n!}{(k-1)!(n-k)!} \quad v^{k-1}(1-v)^{n-k} & 0 \le v \le 1 \\ 0 & \text{ellers}\ . \end{cases}$$

Finn $E(V_k)$ og $\mathrm{Var}(V_k)$.
[Hint: Vis først at $\int_0^1 v^{k-1}(1-v)^{n-k}dv = \frac{(k-1)!(n-k)!}{n!}$]

d) La $X_1, \dots, X_n$ være uavhengige dobbelteksponensielt fordelte observasjoner og la $X_{(1)} \le \dots \le X_{(n)}$ være ordningsobservatoren. Begrunn at et spredningsplott $(F^{-1}(\frac{k}{n+1}), X((k)),\ k = 1, \dots, n$ tilnærmet vil være en rett linje gjennom origo. Hva er stigningsforholdet?

e) La Φ være den kumulative fordelingsfunksjonen til en normalfordelt variabel med forventning 0 og varians 1. Forklar hvordan et plott av $(\Phi^{-1}(\frac{k}{n+1}), X_{(k)}),\ k = 1, \dots, n$ ser ut.

(Fortsettes side 3.)

f) Anta nå at $X_1, \ldots, X_n$ er uavhengige identisk fordelt med tetthet

$$k(x) = \frac{1}{2\sigma} \exp(-|x-\mu|/\sigma)$$

der μ og σ er ukjente parametre. La $X_{(1)} \leq \cdots \leq X_{(k)}$ være ordningsobservatoren.
Hvordan vil nå et plott

$$(F^{-1}\left(\frac{k}{n+1}\right), X_{(k)}) \qquad k = 1, \ldots, n$$

se ut?

Oppgave 3.

Dataene i vedlegg 1 er hentet fra en studie for å undersøke et legemiddel som brukes for å redusere blodtrykket ved operasjoner. Den avhengige variabelen, y (lrec), angir logaritmen av tiden i minutter før blodtrykket er tilbake til det normale etter at behandlingen er avsluttet. De uavhengige variablene er x_1 (ldrug): logaritmen av dosen i milligram og x_2 (bl.-pr): midlere blodtrykk under behandling. I alt 53 personer inngår i studien.

Utskriften fra MINITAB nedenfor angir resultatet av å tilpasse en enkel lineær regresjon der x_1 brukes som uavhengig variabel. Vi antar modellen

$$Y_i = \beta_0 + \beta_1 x_{i1} + \varepsilon_i \qquad i = 1, \ldots, 53$$

der $\varepsilon_1, \ldots, \varepsilon_{53}$ er uavhengige $N(0, \sigma^2)$.

Regression Analysis: lrec versus ldrug

```
The regression equation is
lrec = 1,79 + 0,552 ldrug

Predictor        Coef     SE Coef        T       P
Constant       1,7930      0,5524     3,25   0,002
ldrug          0,5515      0,2734     2,02   0,049

S = 0,6706      R-Sq = 7,4%      R-Sq(adj) = 5,6%
```

a) La $\widehat{\beta}_0$ og $\widehat{\beta}_1$ være minste kvadraters estimatoren til β_0 og β_1. La $S^2 = \frac{1}{51}\sum_{i=1}^{53}(Y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_{i1})^2$. Hvilken fordeling har $\frac{(\widehat{\beta}_1 - \beta_1)}{S}\sqrt{\sum_{i=1}^{53}(x_{i1} - \bar{x}_1)^2}$ der $\bar{x}_1 = \frac{1}{53}\sum_{i=1}^{53} x_{i1}$? Begrunn svaret.

b) Forklar hvordan man kan teste hypotesen $H_0 : \beta_1 = 0$ mot $H_1 : \beta_1 \neq 0$ på grunnlag av utskriften ovenfor. Hvordan kan en teste $H_0 : \beta_1 = 1$ mot $H_1 : \beta_1 \neq 1$? Hva er omtrentlig P-verdi for denne testen?

(Fortsettes side 4.)

c) Finn et 90% konfidensintervall for β_1.

Nedenfor er det angitt en utskrift fra MINITAB der også x_2 er brukt som uavhengig variabel. Nå er modellen

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \eta_i \qquad i = 1, \ldots, 53$$

der $\eta_1, \ldots, \eta_{53}$ er uavhengige $N(0, \sigma^2)$.

Regression Analysis: lrec versus ldrug; bl-pr

```
The regression equation is
lrec = 3,09 + 0,858 ldrug - 0,0288 bl-pr

Predictor        Coef      SE Coef          T        P
Constant       3,0906       0,7973       3,88    0,000
ldrug          0,8579       0,2986       2,87    0,006
bl-pr        -0,02876      0,01314      -2,19    0,033

S = 0,6470      R-Sq = 15,5%     R-Sq(adj) = 12,1%
```

Hvis X er i 53×3 matrisen $\begin{pmatrix} 1 & x_{11} & x_{12} \\ \vdots & & \\ 1 & x_{531} & x_{532} \end{pmatrix}$ er

$$(X'X)^{-1} = \begin{matrix} 1{,}51862 & -0{,}13279 & -0{,}01862 \\ -0{,}13279 & 0{,}21300 & -0{,}00440 \\ -0{,}01862 & -0{,}00440 & 0{,}00041 \end{matrix}.$$

d) Diskuter føyningen av denne modellen i forhold til den forrige.

e) Et rimelig spørsmål er om kovariater av typen $x_3 = x_1 x_2$, d.v.s. produktet av log dose og midlere blodtrykk under behandlingen vil gi noen økt forklaringskraft. Figuren i vedlegg 2 viser et spredningsdiagram mellom y og x_3. Hva er din oppfatning? Gi en kort begrunnelse.

f) Angi et estimat for forventet verdi av tiden det tar før en person har normalt blodtrykk hvis $x_1 = 2.0$ og $x_2 = 60$. Finn også et 90% konfidensintervall for denne størrelsen.

g) Anta nå at en planlegger å gi 5 personer en dose på $x_1 = 2.0$ og at middelverdien av blodtrykket under behandlingen ventes å være lik $x_2 = 60$ for hver enkelt. Angi et 90% prediksjonsintervall for gjennomsnittet av de fem behandlingene.

SLUTT